



# A folding methodology for multivariate extremes: estimation of the spectral probability measure and actuarial applications

Armelle Guillou, Philippe Naveau, Alexandre You

## ► To cite this version:

Armelle Guillou, Philippe Naveau, Alexandre You. A folding methodology for multivariate extremes: estimation of the spectral probability measure and actuarial applications. *Scandinavian Actuarial Journal*, 2015, 2015 (7), 10.1080/03461238.2013.864326 . hal-01312987

**HAL Id: hal-01312987**

**<https://hal.science/hal-01312987>**

Submitted on 11 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A folding methodology for multivariate extremes : estimation of the spectral probability measure and actuarial applications

Armelle Guillo<sup>1</sup>, Philippe Naveau<sup>2</sup>, Alexandre You<sup>3</sup>

<sup>1</sup> IRMA, UMR 7501, Université de Strasbourg et CNRS, France

<sup>2</sup> Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL-CNRS, France

<sup>3</sup> Société Générale Insurance - Sogessur, Direction Technique, France

June 23, 2013

## Abstract

In this paper, the *folding* methodology developed in the context of univariate Extreme Value Theory (EVT) by You et al. (2010) is extended to a multivariate framework. Under the usual EVT assumption of regularly varying tails, our multivariate folding allows for the estimation of the spectral probability measure. A new weakly consistent estimator based on the classical empirical estimator is proposed. Its behaviour is illustrated through simulations and an actuarial application relative to reinsurance pricing in the case of an insurance dataset.

**Key words:** Extreme Value Theory, folding, multivariate regular variation, spectral probability measure, point processes, reinsurance pricing.

## 1 Motivation

In practical applications where extreme phenomena are involved, modeling dependence among extremes is a fundamental issue. Actuarial sciences give rise to many situations where multivariate extremes play a central role. In non life insurance, several lines of business in a portfolio (motor, household, commercial,...) can be exposed to the same catastrophic events (windstorm, floods,...). Non proportional reinsurance programs can be set up in order to cover simultaneously these lines of business against such events. In the context of the Solvency 2 prudential framework, the SCR ("Solvency Capital Requirement"), which is the risk-based regulatory amount to hold against unforeseen extreme events, covers all the potential (and quantifiable) sources of risk (assets, insurance claims, operational,...) and takes the dependence among risks into account (the so-called diversification effect). More generally, multivariate extremes can arise in situations where many variables are involved or where spatial and/or time dependence occurs.

In this paper, we are particularly interested in the estimation of quantities related to rare events in a multivariate framework. Actuaries often have to deal with the computation of complex nonlinear functions of random variables. For bivariate data  $(X, Y)$ , a general expression for these functions of interest is:

$$Q = \mathbb{E}[g(X, Y)] \tag{1}$$

Among examples of such functions, one can cite reinsurance premiums or the probability of a rare event (e.g. attachment point of an excess-of-loss treaty or stress scenario combining univariate claims assumptions). To illustrate the usefulness and practical relevance of our approach, we will use the Loss-ALAE dataset, already analyzed in Frees & Valdez (1998). It consists of a set of  $n = 1500$  general liability claims randomly chosen from late settlement lags and that were provided by Insurance Services Office, Inc. Each claim consists of an indemnity payment (the loss) and an allocated loss adjustment expense (ALAE), that is, an expense specifically attributable to the settlement of the individual claim such as lawyer's fees and claims investigation expenses. For each claim, the total amount paid by the insurance company is equal to the sum of the loss ( $X$ ) and the ALAE ( $Y$ ).

To reach this goal, we develop a methodology based on the estimation of the spectral probability measure, whose definition is recalled in Section 2.1. The estimation method used in this paper is an extension of the folding approach (see You et al., 2010), recalled in Section 2.2, to the multivariate setting (see Section 2.3). A class of estimators of the spectral probability measure based on this folded method is proposed in Section 3 and their weak consistency are also established. The performances of these estimators through simulations are illustrated in Section 4. Finally, also in this last section, we come back to our initial estimation problem of (1) with our set of real insurance data. Proofs are postponed to Section 5.

## 2 Spectral probability measure and folding method

In this section, we start to recall the definition of the spectral measure, afterwards we present the folding methodology developed in You et al. (2010). Then we define the multidimensional version of the folding transformation in a multivariate extreme value framework with some application to statistical estimation problems.

### 2.1 Spectral probability measure

In the following, we consider a random vector  $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$  with a distribution function  $F$  on  $\mathbb{R}_+^d = [0, \infty)^d$ . Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^d$  and  $\mathbb{N} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$  the unit sphere of  $\mathbb{R}^d$  with respect to this norm. Denote by

$$T(\mathbf{X}) := (R, \Theta) = \left( \|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right)$$

the polar coordinates of  $\mathbf{X}$ , where  $R$  is the *radius* of  $\mathbf{X}$  and  $\Theta$  its *polar angle*. Define the spaces  $E := \overline{\mathbb{R}}_+^d \setminus \{\mathbf{0}\}$  and  $\mathbb{N}_+ := \mathbb{N} \cap E$  where  $\overline{\mathbb{R}}_+ := [0, \infty]$ . Note also by  $M_+(E)$  and  $M_+((0, \infty] \times \mathbb{N}_+)$  the sets of all nonzero Radon measures on  $E$  and  $(0, \infty] \times \mathbb{N}_+$ , respectively. We assume that  $\mathbf{X}$  has a regularly varying tail  $\overline{F} := 1 - F$ , that is, there exist a Radon measure  $\nu$  on  $E$  and  $b_n \rightarrow \infty$  such that

$$n \mathbb{P} \left( \frac{\mathbf{X}}{b_n} \in \cdot \right) \xrightarrow{v} \nu(\cdot), \quad n \rightarrow \infty \quad (2)$$

in  $M_+(E)$ , where  $\xrightarrow{v}$  denotes the vague convergence of measures. An equivalent statement in terms of the polar coordinates is the following: there exist a probability measure  $S(\cdot)$  on Borel subsets of  $\mathbb{N}_+$  and  $b_n \rightarrow \infty$  such that

$$n \mathbb{P} \left( \left( \frac{R}{b_n}, \Theta \right) \in \cdot \right) \xrightarrow{v} c \nu_\alpha \times S(\cdot), \quad n \rightarrow \infty \quad (3)$$

in  $M_+((0, \infty] \times \mathbb{N}_+)$  for some  $c > 0$  and  $\alpha > 0$ , where  $\nu_\alpha(r, \infty] = r^{-\alpha}$ . Hence, the radial component and the angular component are asymptotically independent. The so-called *spectral probability measure*  $S$  (de Haan & Resnick, 1977) summarizes the dependence among extremes. In statistics of multivariate extremes, estimation of the spectral probability measure is a crucial

problem which has focused much attention in the literature (see, e.g., Einmahl et al., 1993, 1997, 2001; Einmahl & Segers, 2009).

## 2.2 Univariate folding methodology

Recently, You et al. (2010) introduced a new approach for statistical estimation problems in Extreme Value Theory (EVT). Their method relies on the use of a *folding* transformation, defined by Corcoran & Schneider (2003) in the context of “perfect sampling”. The explicit formulation and the fundamental property of this folding transformation are given in the following lemma.

**Lemma 1.** *Let  $X$  be a random variable with an absolutely continuous distribution function  $F$  and a real number  $u < \tau_F$  where  $\tau_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$  is the right endpoint of  $F$ . We can define the following folding random variable*

$$X^{(F)}(u) = \begin{cases} F^{\leftarrow} \left( \frac{\bar{F}(u)}{F(u)} F(X) + F(u) \right) & \text{if } X < u, \\ X & \text{if } X \geq u, \end{cases} \quad (4)$$

where  $\bar{F} := 1 - F$  is the survival function of  $X$  and  $F^{\leftarrow}$  is the inverse function of  $F$ . Then  $\mathbb{P}(X^{(F)}(u) \leq x) = \mathbb{P}(X \leq x | X > u)$  for all  $x$ .

When  $X$  falls under the fixed threshold  $u$ , the folding transformation defined by (4) can be decomposed into three steps, as illustrated in Figure 1. Firstly, the variable  $X$  is transformed into  $F(X)$ , which takes its values in  $[0, F(u))$ . Secondly, a scaling/shifting transformation is applied in order to obtain a random variable taking its values in  $[F(u), 1)$ . Thirdly, the inverse function  $F^{\leftarrow}$  is applied to bring the modified variable into  $[u, \tau_F)$ .

This transformation has been studied in the univariate independent and identically distributed

Figure 1: The folding mechanism

(iid) case and an adaptation of the classical Peaks-Over-Thresholds approach has been proposed. Simulations tend to indicate that this method improves significantly the performances of extreme quantiles estimators in finite sample situations. Below, we define a multidimensional version of the folding transformation and the statistical procedure which allows us to build the folded version of the sample.

## 2.3 A multivariate folding transformation

The following definition extends the univariate folding transformation and Lemma 2 below its fundamental property studied by You et al. (2010) to a multivariate extreme value framework.

**Definition 1.** *Let  $\mathbf{X}$  be a  $\mathbb{R}_+^d$ -valued random vector with distribution  $F$  and suppose that the tail  $\bar{F} := 1 - F$  is regularly varying. For  $u > 0$ , define the folded random vector  $\mathbf{X}_u^{(F)}$  by its polar coordinates*

$$T(\mathbf{X}_u^{(F)}) = (R_u^{(F)}, \boldsymbol{\Theta}_u^{(F)})$$

where

$$\bullet R_u^{(F)} := F_R^{\leftarrow} \left( \frac{\bar{F}_R(u)}{F_R(u)} F_R(R) + F_R(u) \right) \mathbb{1}\{R \leq u\} + R \mathbb{1}\{R > u\} \quad (5)$$

$$\bullet \boldsymbol{\Theta}_u^{(F)} := \boldsymbol{\Theta}_S \mathbb{1}\{R \leq u\} + \boldsymbol{\Theta} \mathbb{1}\{R > u\} \quad (6)$$

$$\bullet \boldsymbol{\Theta}_S \text{ is a random vector on } \mathbb{N}_+ \text{ that is independent of } R \text{ and such that } \mathbb{P}(\boldsymbol{\Theta}_S \in \cdot) = S(\cdot) \quad (7)$$

with  $F_R(u) = 1 - \bar{F}_R(u) = \mathbb{P}(R \leq u)$ .

**Lemma 2.** Let  $(R_u^{(F)}, \Theta_u^{(F)})$  be defined as in Definition 1. Then, for all  $x \geq 1$ ,

$$\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \cdot) \xrightarrow{v} x^{-\alpha} \times S(\cdot), \text{ as } u \rightarrow \infty$$

in  $M_+(\mathbb{R}_+)$ .

Notice that the random variables  $R_u^{(F)}$  and  $\Theta_u^{(F)}$  are well defined polar coordinates since  $\|\Theta_u^{(F)}\| = \|\Theta_S\| \mathbb{1}\{R \leq u\} + \|\Theta\| \mathbb{1}\{R > u\} = 1$ .

The folding transformation defined in Definition 1 consists in a two-step mechanism. In the first step (“radial folding”), any point whose radius falls below the “circular” threshold  $u$  is sent above it through a univariate folding transformation (see Lemma 1) which is applied to its radius. In the second step (“angular folding”), its polar angle is re-drawn from the spectral probability measure  $S$ , i.e. the angular distribution for extreme points far from the origin.

Now we consider a sample of  $n$  iid random vectors  $\mathbf{X}_1, \dots, \mathbf{X}_n$  with the distribution function  $F$  on  $\mathbb{R}_+^d$ . Let  $T(\mathbf{X}_i) = (R_i, \Theta_i)$ ,  $i = 1, \dots, n$ , be the polar coordinates and  $R_{1,n} \leq \dots \leq R_{n,n}$  the order statistics of the sample radii  $R_1, \dots, R_n$ . To apply the folding transformation defined in Definition 1 to our sample, we have to estimate the unknown quantities: the distribution function  $F_R$  and its inverse  $F_R^{\leftarrow}$  are needed for folding the radii  $R_1, \dots, R_n$  (see (5)), whereas the spectral probability measure  $S$  is required for folding the polar angles  $\Theta_1, \dots, \Theta_n$  (see (6) and (7)). We choose the following preliminary estimators based on the original (i.e. non folded) sample:

- For  $r \leq u$ ,  $F_R(r)$  is estimated by the empirical distribution function

$$F_{n,R}(r) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i \leq r\} \quad (8)$$

whereas  $\bar{F}_R(r)$  is estimated by  $\bar{F}_{n,R}(r) := 1 - F_{n,R}(r)$ .

- Concerning the estimation of the inverse  $F_R^{\leftarrow}$ , for  $0 < p < 1$ , we denote by  $r_p$  the quantile of order  $1-p$  of  $R$ , that is,  $\bar{F}_R(r_p) = \mathbb{P}(R > r_p) = p$ . From the regular variation condition, we get

$$\mathbb{P}(R > r_p) = \frac{\mathbb{P}(R > \frac{r_p}{u} u)}{\mathbb{P}(R > u)} \mathbb{P}(R > u) \approx \left(\frac{r_p}{u}\right)^{-\alpha} \mathbb{P}(R > u).$$

Hence  $r_p \approx u (p/\bar{F}_R(u))^{-1/\alpha}$ . This approximation suggests an estimation for  $r_p = F_R^{\leftarrow}(1-p)$ , provided that  $\bar{F}_R(u)$  and  $\alpha$  are also replaced by some estimators. As above,  $\bar{F}_R(u)$  is estimated by  $\bar{F}_{n,R}(u)$ , whereas the index  $\alpha$  can be estimated by one of the many estimators that are available in the EVT literature (see, e.g., Hill, 1975; Pickands, 1975; Dekkers et al., 1989; Beirlant et al., 1996, 2005). We use the classical Hill (1975) estimator

$$1/\hat{\alpha} := H_{N_u,n} := \frac{1}{N_u} \sum_{j=1}^n \log \frac{R_j}{u} \mathbb{1}\{R_j > u\}$$

with  $N_u := \sum_{j=1}^n \mathbb{1}\{R_j > u\}$ . It follows that an estimator of  $F_R^{\leftarrow}$  is given by

$$\hat{F}_R^{\leftarrow}(1-p) := \hat{r}_p := u \left( \frac{\bar{F}_{n,R}(u)}{p} \right)^{H_{N_u,n}}. \quad (9)$$

If the threshold  $u$  is chosen as some order statistic of the radii  $R_1, \dots, R_n$ , the above estimator is the Weissman (1978) estimator.

- A preliminary estimation of the spectral probability measure  $S$  is given by the empirical spectral probability measure (see, e.g., Einmahl et al., 2001; de Haan & Ferreira, 2006; Resnick, 2006)

$$\widehat{S}(\cdot) := \frac{\sum_{i=1}^n \mathbb{1}\{R_i > R_{n-k,n}, \Theta_i \in \cdot\}}{\sum_{i=1}^n \mathbb{1}\{R_i > R_{n-k,n}\}} \quad (10)$$

where  $k$  is the number of upper order statistics taken into account for estimating  $S$ , or, equivalently, the number of radii larger than the radial threshold  $R_{n-k,n}$ . This estimator is weakly consistent (see Theorem 2 in Section 5).

From (5), the *folded* versions of the radii  $R_1, \dots, R_n$  can be therefore defined by

$$\widehat{R}_{i,u}^{(F)} := \widehat{F}_R^{\leftarrow} \left( \delta(n) \frac{\overline{F}_{n,R}(u)}{F_{n,R}(u)} F_{n,R}(R_i) + F_{n,R}(u) \right) \mathbb{1}\{F_{n,R}(R_i) \leq F_{n,R}(u)\} + R_i \mathbb{1}\{F_{n,R}(R_i) > F_{n,R}(u)\}$$

where  $\delta(n)$  is some continuity correction which is applied in the above expression in order to prevent the case  $\widehat{R}_{i,u}^{(F)} = \infty$  which can occur for  $R_i$  such that  $F_{n,R}(R_i) = F_{n,R}(u)$ . It will be chosen such that  $0 < \delta(n) < 1$  and  $\delta(n) \rightarrow 1$  as  $n \rightarrow \infty$ . Some good candidates are e.g.  $\delta(n) = n/(n+1)$  or  $\delta(n) = (n-1)/n$ . Expression (9) gives

$$\widehat{R}_{i,u}^{(F)} = u \left( 1 - \delta(n) \frac{F_{n,R}(R_i)}{F_{n,R}(u)} \right)^{-H_{N_{u,n}}} \mathbb{1}\{F_{n,R}(R_i) \leq F_{n,R}(u)\} + R_i \mathbb{1}\{F_{n,R}(R_i) > F_{n,R}(u)\} \quad (11)$$

whereas the *folded* versions of the polar angles  $\Theta_1, \dots, \Theta_n$  can be defined by

$$\widehat{\Theta}_{i,u}^{(F)} := \Theta_{\widehat{S},i} \mathbb{1}\{F_{n,R}(R_i) \leq F_{n,R}(u)\} + \Theta_i \mathbb{1}\{F_{n,R}(R_i) > F_{n,R}(u)\} \quad (12)$$

where  $\Theta_{\widehat{S},1}, \dots, \Theta_{\widehat{S},n}$  are iid random vectors on  $\mathbb{N}_+$  such that  $\mathbb{P}(\Theta_{\widehat{S},1} \in \cdot | \mathbf{X}_1, \dots, \mathbf{X}_n) = \widehat{S}(\cdot)$ . Finally, the *folded* versions of the random vectors  $\mathbf{X}_1, \dots, \mathbf{X}_n$  are

$$\widehat{\mathbf{X}}_{i,u}^{(F)} := T^{\leftarrow} \left( \widehat{R}_{i,u}^{(F)}, \widehat{\Theta}_{i,u}^{(F)} \right) = \widehat{R}_{i,u}^{(F)} \widehat{\Theta}_{i,u}^{(F)}, \quad i = 1, \dots, n. \quad (13)$$

A practical example of the folding transformation for a bivariate sample is shown in Figures 2 and 3. The original sample is plotted in Figure 2 whereas the folded sample is displayed in Figure 3 (left panel). In the latter, the radial threshold  $u$  is materialized by the “empty” area around the origin, illustrating the fact that the original sample has been replaced by pseudo-observations (as defined by (13)) that are all above the threshold  $u$ .

### 3 Estimation of the spectral probability measure with the folded sample

Now we can propose a new estimator of the spectral probability measure  $S$  based on the folded sample  $\widehat{\mathbf{X}}_{1,u}^{(F)}, \dots, \widehat{\mathbf{X}}_{n,u}^{(F)}$ . The idea is to replace the observations by their folded versions in the empirical spectral probability measure (10), that is,

$$\widehat{S}_u^{(F)}(\cdot) := \frac{\sum_{i=1}^n \mathbb{1}\left\{\widehat{R}_{i,u}^{(F)} > R_{n-k,n}, \widehat{\Theta}_{i,u}^{(F)} \in \cdot\right\}}{\sum_{i=1}^n \mathbb{1}\left\{\widehat{R}_{i,u}^{(F)} > R_{n-k,n}\right\}}.$$

Given that  $\widehat{R}_{i,u}^{(F)}/R_{n-k,n}$  are actually the relative exceedances above the threshold  $R_{n-k,n}$ , it seems natural to choose  $u = u_{k,n} = R_{n-k,n}$  as the folding threshold. Expressions (11) and (12) which define the folded polar coordinates become respectively

$$\widehat{R}_{i,u_{k,n}}^{(F)} = R_{n-k,n} \left( 1 - \delta(n) \frac{r_{R,i}}{n-k} \right)^{-H_{k,n}} \mathbb{1}\{r_{R,i} \leq n-k\} + R_i \mathbb{1}\{r_{R,i} > n-k\} \quad (14)$$

and

$$\widehat{\Theta}_{i,u_{k,n}}^{(F)} = \Theta_{\widehat{S},i} \mathbb{1}\{r_{R,i} \leq n-k\} + \Theta_i \mathbb{1}\{r_{R,i} > n-k\}, \quad i = 1, \dots, n \quad (15)$$

where  $r_{R,i}$ ,  $i = 1, \dots, n$ , are the rank statistics of the radii  $R_1, \dots, R_n$ , that is,  $r_{R,i} := \sum_{j=1}^n \mathbb{1}\{R_j \leq R_i\}$ ,  $i = 1, \dots, n$ . Our *folded* estimator of the spectral probability measure  $S$  can be rewritten as

$$\begin{aligned} \widehat{S}_{u_{k,n}}^{(F)}(\cdot) &:= \frac{\sum_{i=1}^n \mathbb{1}\{\widehat{R}_{i,u_{k,n}}^{(F)} > u_{k,n}, \widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\}}{\sum_{i=1}^n \mathbb{1}\{\widehat{R}_{i,u_{k,n}}^{(F)} > u_{k,n}\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\} \end{aligned} \quad (16)$$

since  $\widehat{R}_{i,u_{k,n}}^{(F)} > u_{k,n}$  for all  $i = 1, \dots, n$  by construction. Our estimator is then nothing else than the empirical distribution function of the folded polar angles  $\widehat{\Theta}_{1,u_{k,n}}^{(F)}, \dots, \widehat{\Theta}_{n,u_{k,n}}^{(F)}$ . This is a remarkable feature since the folded radii  $\widehat{R}_{1,u_{k,n}}^{(F)}, \dots, \widehat{R}_{n,u_{k,n}}^{(F)}$  do not enter into the estimation of the spectral probability measure (i.e. it is only required that all the folded radii overshoot the threshold).

The next result gives the weak consistency of the folded estimator of  $S$ .

**Proposition 1.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid random vectors with a distribution function  $F$  on  $\mathbb{R}_+^d$  whose tail is regularly varying and with polar coordinates  $T(\mathbf{X}_i) = (R_i, \Theta_i)$ ,  $i = 1, \dots, n$ . Denote by  $(\widehat{R}_{i,u_{k,n}}^{(F)}, \widehat{\Theta}_{i,u_{k,n}}^{(F)})$ ,  $i = 1, \dots, n$ , the folded polar coordinates as defined by (14) and (15), respectively, with  $u_{k,n} = R_{n-k,n}$ . Then, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , we have*

$$\widehat{S}_{u_{k,n}}^{(F)}(\cdot) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\} \xrightarrow{d} S(\cdot) \quad (17)$$

in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

The previous statistical framework is valid under the regular variation assumption (2). As usual, we assume that all the marginals of  $\mathbf{X}$  are tail equivalent to a Pareto distribution with the same extreme value index  $\alpha$ . This corresponds to the so-called *standard case* of regular variation where  $b_n = n$  (or  $b(t) = t$ ) and  $\alpha = 1$  (de Haan & Resnick, 1977; Resnick, 1987). In practical situations, marginals of a random vector are generally not tail equivalent and, though it is convenient for theoretical developments, the preceding framework is a bit restrictive for applications. Some techniques have been proposed in the literature in order to get back to the standard case when starting from more generic (nonstandard) regular variation conditions. We propose here to use the ranks method (see, e.g., de Haan & de Ronde, 1998; Einmahl et al., 2001; de Haan & Ferreira, 2006; Resnick, 2006); see Theorem 1 in Section 5 for the theoretical motivation of the method. The underlying idea is to transform the marginals of the random vector into unit Pareto distributions via a nonparametric approach: Denoting by  $X_i^{(j)}$  the  $j$ -th component of  $\mathbf{X}_i$ , the rank transformed sample is defined by

$$\left( \frac{n}{n+1-r_{X^{(j)},i}}, j = 1, \dots, d \right), \quad i = 1, \dots, n,$$

where  $r_{X^{(j)},i}$ ,  $i = 1, \dots, n$ , are the rank statistics of the univariate sample  $X_1^{(j)}, \dots, X_n^{(j)}$ , that is,  $r_{X^{(j)},i} := \sum_{m=1}^n \mathbb{1}\{X_m^{(j)} \leq X_i^{(j)}\}$ . Taking the polar coordinates

$$T\left(\frac{n}{n+1-\mathbf{r}_i}\right) =: (R_i, \Theta_i), \quad i = 1, \dots, n,$$

where  $\mathbf{r}_i := (r_{X^{(1)},i}, \dots, r_{X^{(d)},i})$ , the folding can be applied according to formulæ (11) and (12) with the folding threshold chosen as  $u_{k,n} = n/k$ . We obtain the folded polar coordinates  $(\widehat{R}_{i,u_{k,n}}^{(F)}, \widehat{\Theta}_{i,u_{k,n}}^{(F)})$ ,  $i = 1, \dots, n$ , and the folded empirical estimator (16) can be computed. Notice that, as a consequence of the rank transformation, the estimator is invariant under monotone transformations of the marginals. In particular, it does not depend on the marginal distributions of the random vector  $\mathbf{X}$ . The next result states that the folded empirical spectral probability measure based on the rank transformed data is weakly consistent.

**Proposition 2.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid random vectors with a distribution function  $F$  on  $\mathbb{R}_+^d$  whose tail satisfies the nonstandard regular variation conditions (23) and (24). The polar coordinates of the rank transformed observations are denoted by  $T\left(\frac{n}{n+1-\mathbf{r}_i}\right) = (R_i, \Theta_i)$ ,  $i = 1, \dots, n$ . Let  $(\widehat{R}_{i,u_{k,n}}^{(F)}, \widehat{\Theta}_{i,u_{k,n}}^{(F)})$ ,  $i = 1, \dots, n$ , be the folded polar coordinates, with  $u_{k,n} = n/k$ . Then, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , we have*

$$\widehat{S}_{u_{k,n}}^{(F)}(\cdot) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\}} \xrightarrow{d} S(\cdot) \quad (18)$$

in  $M_+(\mathbb{R}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

## 4 Simulations and an insurance data application

An important step on the way of estimating quantities of interest related to multivariate extremes such as defined by (1) is the determination of the spectral probability measure. To this aim, we start to illustrate the efficiency of our approach in estimating the spectral probability measure on two simulated examples: the autoregressive conditional heteroscedasticity (ARCH) model, used in financial time series modelling, and the asymmetric logistic model, used in copula modelling. Then, we come back to our initial problem of estimating expressions of the form (1) and propose a simple approach based on Monte Carlo simulations. An application of this simulation-based approach is shown by estimating reinsurance premiums with the Loss-ALAE dataset.

### 4.1 ARCH(1) process

We consider the ARCH(1) process defined by

$$X_t = (\beta + \lambda X_{t-1}^2)^{1/2} Z_t, \quad t \in \mathbb{N} \quad (19)$$

where  $(Z_t)$  is a sequence of iid standard normal random variables and  $\beta, \lambda > 0$ . ARCH processes have been introduced by Engle (1982) and are widely used for financial time series modelling (e.g. exchange rates), especially when one has to deal with volatility clustering. The mathematical properties of such processes have been extensively studied in the literature (see, e.g., Embrechts et al., 1997, Section 8.4). In particular, it is well-known that equation (19) admits a stationary solution  $(X_t)$  if  $0 < \lambda < 2e^\gamma$ , where  $\gamma \approx 0.5772$  is the Euler's constant. This solution has infinite variance if  $1 \leq \lambda < 2e^\gamma$  and although the innovations  $(Z_t)$  are light-tailed, the tail behavior of  $X_t$  is of Pareto-type:

$$\mathbb{P}(X_t > x) \sim c_{\beta,\lambda,\kappa} x^{-\alpha}, \quad x \rightarrow \infty$$

where  $\alpha$  is the unique solution of the equation  $\mathbb{E}(\lambda Z^2)^{\alpha/2} = 1$  with  $Z \sim \mathcal{N}(0, 1)$  and  $c_{\beta,\lambda,\alpha}$  is a constant that depends on  $\beta$ ,  $\lambda$  and  $\alpha$ . The equation  $\mathbb{E}(\lambda Z^2)^{\alpha/2} = 1$  must be solved using numerical methods.

If we consider now the pair  $(X_t, X_{t+1})$  with  $(X_t)$  a stationary ARCH(1) process, the distribution tail of this random vector is regularly varying (see Davis & Mikosch, 1998, Lemma A.1), that is



there exists a sequence  $b_n \rightarrow \infty$  such that

$$\begin{aligned} & n \mathbb{P} \left( b_n^{-1} \|(X_t, X_{t+1})\| > r, \frac{(X_t, X_{t+1})}{\|(X_t, X_{t+1})\|} \in \cdot \right) \\ & \xrightarrow{d} r^{-\alpha} \frac{\mathbb{E} \left( \|(B_t, \lambda^{1/2} Z_{t+1})\|^\alpha \mathbb{1} \left\{ \frac{(B_t, \lambda^{1/2} Z_{t+1})}{\|(B_t, \lambda^{1/2} Z_{t+1})\|} \in \cdot \right\} \right)}{\mathbb{E} (\|(B_t, \lambda^{1/2} Z_{t+1})\|^\alpha)} \end{aligned} \quad (20)$$

in  $M_+(\mathbb{N} \cap E)$  as  $n \rightarrow \infty$ , where  $E = \mathbb{R}^2 \setminus \{\mathbf{0}\}$  and  $(B_t)$  is a sequence of iid random variables defined by  $\mathbb{P}(B_t = 1) = \mathbb{P}(B_t = -1) = 1/2$  and independent of  $|X_t|$ . The chosen norm is the  $L_2$ -norm, i.e.  $\|\mathbf{x}\|^2 = (x^{(1)})^2 + (x^{(2)})^2$ .

Let us take the ARCH(1) process with parameters  $(\beta, \lambda) = (1, 0.9)$ . A numerical routine gives us  $\alpha = 2.3$ . We want to estimate the spectral probability measure in (20). To illustrate our folding procedure in this case,  $n = 5000$  values of  $(X_t)$  have been generated from the ARCH(1) process (see Figure 2). The folded versions of  $(X_t, X_{t+1})$  have been obtained over a threshold  $u_{k,n}$ , taken as the empirical quantile of the radii  $R_t := \|(X_t, X_{t+1})\|$  at level 0.95 (i.e.  $k = 250$  exceedances).

Figure 3 (left) represents the scatterplot of the folded versions of  $(X_t, X_{t+1})$  with the “true” exceedances (i.e. exceedances before application of the folding method) together with the “pseudo-exceedances” (i.e. observations initially below the threshold). Also Figure 3 (right) illustrates the comparison between the spectral density estimates using the conventional estimator (10), the folded estimator (16) and the true spectral density which has been computed numerically. The folding yields a density estimate that seems closer to the true spectral density than the conventional empirical estimator. This result is confirmed with a Mean Integrated Squared Error criterion (MISE) based on 1000 sequences of length  $n = 5000$  from the ARCH(1) process and computed for several values of the thresholding parameter  $k$  (from 50 to 2000). For these two methods, the empirical MISE have been computed over the 1000 samples as follows:

$$\widehat{MISE}(\widehat{S}) := \frac{1}{1000} \sum_{i=1}^{1000} \int_0^{2\pi} \left( \widehat{S}_i([0, \theta]) - S([0, \theta]) \right)^2 d\theta \quad (21)$$

$$\widehat{MISE}(\widehat{S}_{u_{k,n}}^{(F)}) := \frac{1}{1000} \sum_{i=1}^{1000} \int_0^{2\pi} \left( \widehat{S}_{u_{k,n},i}^{(F)}([0, \theta]) - S([0, \theta]) \right)^2 d\theta \quad (22)$$

where  $S$ ,  $\widehat{S}_i$  and  $\widehat{S}_{u_{k,n},i}^{(F)}$  denote the true spectral measure from (20), the conventional estimator and the folded estimator for the  $i$ -th sequence, respectively. Results are shown in Figure 4. The folding estimator outperforms the conventional empirical estimator for the whole range of values of the thresholding parameter  $k$ .

Figure 2: Time series plot of the ARCH(1) process (left) and scatterplot of  $(X_t, X_{t+1})$  (right).

## 4.2 Asymmetric logistic model

We consider the bivariate extreme value distribution with unit Fréchet margins, defined by

$$G_0(x, y) := \exp \left\{ -\ell \left( \frac{1}{x}, \frac{1}{y} \right) \right\}, \quad x, y > 0$$

where the function  $\ell$  is called the stable tail dependence function. An example of parametric form for the function  $\ell$  is given by the asymmetric logistic model (Tawn, 1988)

$$\ell(x, y) := (1 - \psi_1)x + (1 - \psi_2)y + \{(\psi_1 x)^r + (\psi_2 y)^r\}^{1/r}, \quad x, y \geq 0$$

Figure 3: Scatterplot of the folded versions of  $(X_t, X_{t+1})$  (left): “true” exceedances (“+”) and pseudo-exceedances (dots) over the threshold  $u$ . Spectral density estimates for  $(X_t, X_{t+1})$  (right): conventional empirical estimator (dotted line) and folded estimator (solid line), compared to the true density (grey). To ease graphical representation, density curves have been smoothed using a kernel estimator.

Figure 4: MISE of the conventional empirical estimator (dotted line) and the folded estimator (solid line) for 1000 sequences of length  $n = 5000$  from the ARCH(1) process.

with  $0 \leq \psi_1, \psi_2 \leq 1$  and  $r \geq 1$  (notation:  $\text{LM}(\psi_1, \psi_2, r)$ ). The asymmetric logistic model has been extensively used in the extreme value copulas framework for large claims modelling and excess-of-loss reinsurance issues (see, e.g., Beirlant et al., 2004; Cebrián et al., 2003; Frees & Valdez, 1998). It is a very flexible dependence model which encompasses a wide range of situations, from asymptotic independence ( $r = 1$  or  $\psi_1\psi_2 = 0$ ) to full dependence ( $\psi_1 = \psi_2 = 1$  and  $r = \infty$ ). Values of  $\psi_1$  and  $\psi_2$  different from 1 allow us to build dependence structures whose spectral measure has atoms on  $\pi/2$  and 0, respectively. The special case  $\psi_1 = \psi_2 = 1$  yields the Gumbel’s model (Gumbel, 1960). The tail of  $G_0$  is regularly varying with the limit measure  $\nu$  given by

$$\nu((x, \infty] \times (y, \infty]) = \ell\left(\frac{1}{x}, \frac{1}{y}\right).$$

The spectral probability measure  $S$  on  $[0, \pi/2]$  is given by

$$\begin{cases} S(\{0\}) = \frac{1 - \psi_2}{\int_0^{\pi/2} dQ(\theta)} \\ S(\{\pi/2\}) = \frac{1 - \psi_1}{\int_0^{\pi/2} dQ(\theta)} \\ S(\Lambda) = \frac{\int_\Lambda dQ(\theta)}{\int_0^{\pi/2} dQ(\theta)}, \quad \Lambda \cap \{0, \pi/2\} = \emptyset \end{cases}$$

with

$$\frac{dQ}{d\theta}(\theta) := (r-1)(\psi_1\psi_2)^r (\sin\theta + \cos\theta)^{r-2} (\sin\theta \cos\theta)^{r-2} \{(\psi_1 \cos\theta)^r + (\psi_2 \sin\theta)^r\}^{1/r-2}$$

for  $1 < r < \infty$ .

For this example, we have simulated 1000 samples of size  $n = 1000$  from two logistic models  $\text{LM}(1,1,1.25)$  and  $\text{LM}(1,1,2.5)$ . The chosen norm is again the  $L_2$ -norm. Folding has been performed on the rank transformed samples and the spectral probability measure has been estimated with the three following estimators: the conventional and the folded estimators, as defined by (10) and (16), respectively, and the maximum empirical likelihood estimator (MELE) recently proposed by Einmahl & Segers (2009). The thresholding parameter  $k$  has been allowed to vary. For each model, a scatterplot for one sample is displayed in Figure 5. For  $k = 100$  in the  $\text{LM}(1,1,1.25)$  model and  $k = 200$  in the  $\text{LM}(1,1,2.5)$  model, the folded versions of the rank transformed samples and the spectral density estimates are displayed in Figures 6 and 7, respectively. To ease visualization, Figures 5 and 6 are on a logarithmic scale.

For these three methods, the empirical MISE have been computed over the 1000 samples using

formula (21) and (22) for  $\hat{S}$  and  $\hat{S}_{u_{k,n}}^{(F)}$  and a similar expression for the MELE:

$$\widehat{MISE}(\hat{S}^{MELE}) := \frac{1}{1000} \sum_{i=1}^{1000} \int_0^{\pi/2} \left( \hat{S}_i^{MELE}([0, \theta]) - S([0, \theta]) \right)^2 d\theta$$

where  $\hat{S}_i^{MELE}$  denotes the MELE for the  $i$ -th sample. Results are shown in Figure 8. For both logistic models, the folded estimator is better than both the conventional empirical estimator and the MELE for a wide range of values of  $k$  (the conventional estimator and the MELE perform slightly better for small values of  $k$ ). Moreover, the MISE of the folded estimator seems more stable as  $k$  varies than the conventional estimator. From a practical point of view, this means that making a “wrong” choice of the thresholding parameter  $k$  has a smaller consequence on the statistical performance of the estimation in the case of the folding.

Figure 5: Scatterplots of  $n = 1000$  random vectors from the models LM(1,1,1.25) and LM(1,1,2.5) (on a logarithmic scale).

Figure 6: Folded versions of the samples displayed in Figure 5, after rank transformation (on a logarithmic scale): “true” exceedances (“+”) and pseudo-exceedances (dots) over the thresholds  $u_{k,n} = n/k$ .

Figure 7: Spectral density estimates for the samples displayed in Figure 5: conventional empirical estimator (dotted line), folded estimator (solid line) and MELE (dashed line), compared to the true density (grey). To ease graphical representation, density curves have been smoothed using a kernel estimator.

### 4.3 Application to the Loss-ALAE dataset

We come back to our initial estimation problem presented in Section 1. This dataset of size  $n = 1500$  has been first analyzed in Frees & Valdez (1998) in the context of copula fitting. Spectral measure estimation has been considered in Einmahl & Segers (2009); see also Chapter 9 in Beirlant et al. (2004) for an extended statistical analysis. The data are displayed in Figure 9. The plot suggests a strong relationship between the two components at intermediate levels (the correlation between losses and ALAE is 0.44 in log-scale). Firstly, we seek to estimate the underlying spectral probability measure. In Beirlant et al. (2004), Generalized Pareto distributions are fitted to the marginals conditionally on being above a high threshold, with Pareto index estimates equal respectively to  $\hat{\alpha}_1 = 1.92$  (Loss) and  $\hat{\alpha}_2 = 2.13$  (ALAE). Considering these values, one can assume the equality of tail indices in the following. Hence, the spectral probability measure can be estimated without any prior transformation of the marginals. The following values for the thresholding parameter have been used:  $k = 25, 50, 100, 200$ . The spectral probability measure is estimated using the conventional estimator (10), the folded estimator (16) and the MELE from Einmahl & Segers (2009). Spectral density estimates are displayed in Figure 10. The MELE is basically close to the conventional estimate. For every value of  $k$ , all the estimates indicate a bimodal density with modes located near 0 and  $\pi/2$ , so there is some evidence for

Figure 8: MISE of the conventional empirical estimator (dotted line), the folded estimator (solid line) and the MELE (dashed line) for 1000 samples of size  $n = 1000$  from the models LM(1,1,1.25) and LM(1,1,2.5).

some asymptotic dependence between losses and ALAE, even if it seems that the mass is getting more concentrated around the modes as  $k$  decreases.

With the spectral probability measure being estimated, we can examine some estimation problems related to rare events in a multivariate framework, such as functions  $Q$  defined in (1). In Frees & Valdez (1998), the authors propose to evaluate reinsurance premiums for an excess-of-loss treaty with deductible  $D$  and limit  $L$ . The corresponding function  $g$  in (1) is given by

$$g(X, Y) = \begin{cases} 0 & \text{if } X < D, \\ X - D + \frac{X-D}{X}Y & \text{if } D \leq X < L, \\ L - D + \frac{L-D}{L}Y & \text{if } X \leq L. \end{cases}$$

With such a kind of treaty, each claim triggers the reinsurance cover according to the amount of loss ( $X$ ) and the ALAE ( $Y$ ) are ceded to the reinsurer on a prorata basis. Other possible applications include the estimation of a probability of a rare event (e.g. attachment point of an excess-of-loss treaty or stress scenario combining univariate claims assumptions) with a function  $g$  of the form  $g = \mathbb{1}\{(X, Y) \in B\}$  where  $B$  is a risk region of interest.

These quantities can be estimated using a simple approach based on Monte Carlo simulations:

1. Simulate  $m$  iid pairs  $(U_i, V_i)$ ,  $i = 1, \dots, m$ , with  $U_i$  and  $V_i$  independent uniform  $(0, 1)$  random numbers;
2. Compute  $\tilde{R}_i = u_{k,n}(1 - U_i)^{-1/\hat{\alpha}}$  and  $\tilde{\Theta}_i = \hat{S}_{u_{k,n}}^{(F)\leftarrow}(V_i)$ ,  $i = 1, \dots, m$ , where  $\hat{\alpha}$  is an estimate of the Pareto index of the radial component in (3) and  $\hat{S}_{u_{k,n}}^{(F)\leftarrow}$  is the empirical quantile function for the folded estimate  $\hat{S}_{u_{k,n}}^{(F)}$  of the spectral probability measure  $S$ ;
3. Compute  $(\tilde{X}_i, \tilde{Y}_i) := (\tilde{R}_i \cos \tilde{\Theta}_i, \tilde{R}_i \sin \tilde{\Theta}_i)$ ,  $i = 1, \dots, m$ ;
4. Estimate  $Q = \mathbb{E}[g(X, Y)]$  by

$$\hat{Q} = \frac{1}{m} \sum_{i=1}^m g(\tilde{X}_i, \tilde{Y}_i) \cdot \frac{k}{n} + \frac{\sum_j^n g(X_j, Y_j) \mathbb{1}\{R_j \leq u_{k,n}\}}{\sum_j^n \mathbb{1}\{R_j \leq u_{k,n}\}} \cdot \left(1 - \frac{k}{n}\right)$$

with the expression for  $\hat{Q}$  coming from the fact that

$$\mathbb{E}[g(X, Y)] = \mathbb{E}[g(X, Y)|R > u_{k,n}] \mathbb{P}(R > u_{k,n}) + \mathbb{E}[g(X, Y)|R \leq u_{k,n}] \mathbb{P}(R \leq u_{k,n}).$$

Notice that this simulation program should be slightly adapted in the case of marginals with different tails. Indeed, the simulated outcomes from the folded multivariate distribution are tail equivalent and should be correctly rescaled according to the tail behaviour of each marginal.

Now, we choose to estimate the preceding reinsurance premiums for the Loss-ALAE data. We use our Monte Carlo method with a number of simulated pairs equal to  $m = 500\,000$ . The Pareto index in Step 2 is estimated with a folded Hill estimator, using the univariate folding

approach described in You et al. (2010). The corresponding Hill plot (not shown) is quite stable for a wide range of values of the thresholding parameter  $k$  (Pareto index estimate  $\hat{\alpha} = 1.65$ ). Here, we give our results for  $k = 100$ :

- $D = 50\,000$ ,  $L = 100\,000$ :  $\hat{Q} = 7\,634$  (RoL = 15.27%)
- $D = 75\,000$ ,  $L = 100\,000$ :  $\hat{Q} = 3\,593$  (RoL = 14.37%)
- $D = 95\,000$ ,  $L = 100\,000$ :  $\hat{Q} = 690$  (RoL = 13.80%)
- $D = 500\,000$ ,  $L = 1\,000\,000$ :  $\hat{Q} = 2\,795$  (RoL = 0.56%)
- $D = 750\,000$ ,  $L = 1\,000\,000$ :  $\hat{Q} = 1\,114$  (RoL = 0.45%)
- $D = 950\,000$ ,  $L = 1\,000\,000$ :  $\hat{Q} = 197$  (RoL = 0.39%)

These results show two common features for non proportional reinsurance treaties : firstly, the thinner the layer ( $L - D$ ) the smaller the ceded amount of claims to the reinsurer (and hence the premium of the layer); secondly, the so-called Rate on Line (RoL), which is the price of one monetary unit of reinsurance capacity (i.e. the premium of the layer divided by the size  $L - D$  of the layer) decreases when the deductible increases, that is, upper layers are less costly than lower layers.

Figure 9: Scatterplot of the Loss-ALAE data (on a logarithmic scale).

Figure 10: Spectral density estimates for the Loss-ALAE data: conventional empirical estimator (dotted line), folded estimator (solid line) and MELE (dashed line). To ease graphical representation, density curves have been smoothed using a kernel estimator.

## 5 Appendix

**Theorem 1** (Resnick, 2006, Theorem 6.5, p. 204-205). *Let  $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$  be a random vector on  $\mathbb{R}_+^d$ . Suppose for  $1 \leq j \leq d$  that there exist sequences  $\{b_n^{(j)}, n \geq 1\}$ , with  $\lim_{n \rightarrow \infty} b_n^{(j)} = \infty$ ,  $j = 1, \dots, d$ , such that we have the following:*

(i) *Marginal regular variation: for each  $j = 1, \dots, d$ ,*

$$n \mathbb{P} \left( \frac{X^{(j)}}{b_n^{(j)}} \in \cdot \right) \xrightarrow{v} \nu_{\alpha_j}(\cdot), \quad n \rightarrow \infty \quad (23)$$

*in  $M_+(0, \infty]$ , where  $\nu_{\alpha_j}(x, \infty] = x^{-\alpha_j}$  for  $x > 0$  and  $\alpha_j > 0$ .*

(ii) *Nonstandard global regular variation: there exists a measure  $\nu$  on Borel subsets of  $\mathbb{R}_+^d$  such that*

$$n \mathbb{P} \left( \left( \frac{X^{(j)}}{b_n^{(j)}}, j = 1, \dots, d \right) \in \cdot \right) \xrightarrow{\nu} \nu(\cdot), \quad n \rightarrow \infty \quad (24)$$

*in  $M_+(\mathbb{R}_+^d)$ .*

*Let  $\bar{F}^{(j)}(x) := \mathbb{P}(X^{(j)} > x)$  be the  $j$ -th marginal distribution tail, and from (23), we can define*

$$b^{(j)}(x) := \left( \frac{1}{1 - \bar{F}^{(j)}(\cdot)} \right)^{\leftarrow}(x), \quad x > 1,$$

and set  $b_n^{(j)} = b^{(j)}(n)$ . Then we have the following:

(iii) *Standard global regular variation:*

$$n \mathbb{P} \left( \left( \frac{(b^{(j)})^\leftarrow(X^{(j)})}{n}, j = 1, \dots, d \right) \in \cdot \right) \xrightarrow{v} \nu_*(\cdot), \quad n \rightarrow \infty \quad (25)$$

in  $M_+(\mathbb{R}_+^d)$ , where

$$\nu_*(t \cdot) := t^{-1} \nu_*(\cdot) \quad (26)$$

on Borel subsets of  $\mathbb{R}_+^d$ .

(iv) *Standard marginal convergence:* for  $j = 1, \dots, d$ ,

$$n \mathbb{P} \left( \frac{(b^{(j)})^\leftarrow(X^{(j)})}{n} > x \right) \rightarrow x^{-1}, \quad x > 0, \quad n \rightarrow \infty. \quad (27)$$

The following theorem summarizes some useful results about the weak consistency of empirical spectral probability measures.

**Theorem 2.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid random vectors with a distribution function  $F$  on  $\mathbb{R}_+^d$  whose tail is regularly varying and with polar coordinates  $T(\mathbf{X}_i) = (R_i, \boldsymbol{\Theta}_i)$ ,  $i = 1, \dots, n$ . Suppose the regular variation condition is given by (3). Set  $\tilde{b}(n/k) := c^{1/\alpha} b(n/k)$ . Then*

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\{(R_i/\tilde{b}(n/k), \boldsymbol{\Theta}_i) \in \cdot\} \xrightarrow{d} \nu_\alpha \times S(\cdot) \quad (28)$$

in  $M_+((0, \infty] \times \mathfrak{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

Moreover, if we estimate  $\tilde{b}(n/k)$  by  $R_{n-k,n}$  the  $(k+1)$ -th largest observation among  $R_1, \dots, R_n$ , then

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\{(R_i/R_{n-k,n}, \boldsymbol{\Theta}_i) \in \cdot\} \xrightarrow{d} \nu_\alpha \times S(\cdot) \quad (29)$$

in  $M_+((0, \infty] \times \mathfrak{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ . In particular, we have

$$\hat{S}(\cdot) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{R_i > R_{n-k,n}, \boldsymbol{\Theta}_i \in \cdot\} \xrightarrow{d} S(\cdot). \quad (30)$$

*Proof:*

Result (28) comes directly from Theorem 6.2. in Resnick (2006) (p. 179-180) and from the homogeneity property of  $\nu_\alpha$ . We give below the proof for (29). Then (30) is trivial.

We have

$$\frac{R_{n-k,n}}{\tilde{b}(n/k)} \xrightarrow{\mathbb{P}} 1 \quad (31)$$

as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ . Indeed, for  $\epsilon > 0$ , we have

$$\begin{aligned} \mathbb{P} \left( \left| \frac{R_{n-k,n}}{\tilde{b}(n/k)} - 1 \right| > \epsilon \right) &= \mathbb{P} \left( \frac{R_{n-k,n}}{\tilde{b}(n/k)} > 1 + \epsilon \right) + \mathbb{P} \left( \frac{R_{n-k,n}}{\tilde{b}(n/k)} < 1 - \epsilon \right) \\ &= \mathbb{P} \left( \sum_{i=1}^n \mathbb{1}\{R_i/\tilde{b}(n/k) > 1 + \epsilon\} > k \right) \\ &\quad + \mathbb{P} \left( \sum_{i=1}^n \mathbb{1}\{R_i/\tilde{b}(n/k) < 1 - \epsilon\} \leq k \right). \end{aligned}$$

From (28), as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ , we get

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\left\{R_i/\tilde{b}(n/k) > 1 + \epsilon\right\} \xrightarrow{\mathbb{P}} (1 + \epsilon)^{-\alpha} < 1$$

and

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\left\{R_i/\tilde{b}(n/k) > 1 - \epsilon\right\} \xrightarrow{\mathbb{P}} (1 - \epsilon)^{-\alpha} > 1,$$

which implies

$$\mathbb{P}\left(\left|\frac{R_{n-k,n}}{\tilde{b}(n/k)} - 1\right| > \epsilon\right) \rightarrow 0.$$

Now, set

$$\phi_n(\cdot) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}\left\{(R_i/\tilde{b}(n/k), \Theta_i) \in \cdot\right\} \quad \text{and} \quad \phi(\cdot) := \nu_\alpha \times S(\cdot).$$

From (28) and (31), we have

$$\left(\phi_n, \frac{R_{n-k,n}}{\tilde{b}(n/k)}\right) \xrightarrow{d} (\phi, 1)$$

in  $M_+((0, \infty] \times \mathbb{N}_+) \times (0, \infty)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

Set

$$\hat{\phi}_n(\cdot) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{(R_i/R_{n-k,n}, \Theta_i) \in \cdot\}.$$

The map

$$C : M_+((0, \infty] \times \mathbb{N}_+) \times (0, \infty) \mapsto M_+((0, \infty] \times \mathbb{N}_+)$$

defined by

$$C(\mu, x)(\cdot) = \mu(x \cdot)$$

is continuous at  $(\phi, 1)$ . Since

$$\hat{\phi}_n(\cdot) = \phi_n\left(\frac{R_{n-k,n}}{\tilde{b}(n/k)} \cdot\right) = C\left(\phi_n, \frac{R_{n-k,n}}{\tilde{b}(n/k)}\right)(\cdot),$$

result (29) then follows from the continuous mapping theorem (see, e.g., Theorem 2.7 in Billingsley, 1999, p. 21).

## 5.1 Proof of Lemma 2

Let  $\Lambda$  be a relatively compact Borel subset of  $\mathbb{N}_+$  such that  $S(\partial\Lambda) = 0$ . Clearly,

$$\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda) = \mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda, R \leq u) + \mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda, R > u).$$

Now remark that, for  $x \geq 1$

$$\begin{aligned}
\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda, R \leq u) &= \mathbb{P}\left(F_R^{\leftarrow}\left(\frac{\bar{F}_R(u)}{F_R(u)}F_R(R) + F_R(u)\right) > xu, \Theta_S \in \Lambda, R \leq u\right) \\
&= \mathbb{P}\left(F_R^{\leftarrow}\left(\frac{\bar{F}_R(u)}{F_R(u)}F_R(R) + F_R(u)\right) > xu \mid R \leq u\right) \mathbb{P}(R \leq u) S(\Lambda) \\
&= \mathbb{P}\left(F_R(R) > \frac{F_R(u)}{\bar{F}_R(u)}(F_R(xu) - F_R(u)) \mid F_R(R) \leq F_R(u)\right) \\
&\quad \times \mathbb{P}(R \leq u) S(\Lambda) \\
&= \frac{F_R(u) - \frac{F_R(u)}{\bar{F}_R(u)}(F_R(xu) - F_R(u))}{F_R(u)} \mathbb{P}(R \leq u) S(\Lambda) \\
&= \frac{\bar{F}_R(xu)}{\bar{F}_R(u)} \mathbb{P}(R \leq u) S(\Lambda) \\
&= \mathbb{P}(R > xu \mid R > u) \mathbb{P}(R \leq u) S(\Lambda)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda, R > u) &= \mathbb{P}(R > xu, \Theta \in \Lambda, R > u) \\
&= \mathbb{P}(R > xu, \Theta \in \Lambda \mid R > u) \bar{F}_R(u).
\end{aligned}$$

Therefore

$$\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda) = \mathbb{P}(R > xu \mid R > u) F_R(u) S(\Lambda) + \mathbb{P}(R > xu, \Theta \in \Lambda \mid R > u) \bar{F}_R(u).$$

Recall that (3) is the same as

$$\frac{\mathbb{P}(R > xu, \Theta \in \cdot)}{\mathbb{P}(R > u)} \xrightarrow{v} x^{-\alpha} \times S(\cdot), \quad x > 0 \quad u \rightarrow \infty.$$

It follows that

- $\mathbb{P}(R > xu \mid R > u) \rightarrow x^{-\alpha}$
- $\mathbb{P}(R > xu, \Theta \in \Lambda \mid R > u) \rightarrow x^{-\alpha} \times S(\Lambda).$

Therefore

$$\mathbb{P}(R_u^{(F)} > xu, \Theta_u^{(F)} \in \Lambda) \rightarrow x^{-\alpha} \times S(\Lambda)$$

as  $u \rightarrow \infty$ .

## 5.2 Proof of Proposition 1

We have

$$\begin{aligned}
\widehat{S}_{u_{k,n}}^{(F)}(\cdot) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\} \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i \leq u_{k,n}, \widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\} + \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i > u_{k,n}, \widehat{\Theta}_{i,u_{k,n}}^{(F)} \in \cdot\} \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i \leq u_{k,n}, \Theta_{\widehat{S},i} \in \cdot\} + \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i > u_{k,n}, \Theta_i \in \cdot\} \\
&=: A + B.
\end{aligned}$$

We study each term separately.



**Study of the term  $A$ :**

Let  $R_{i_1}, \dots, R_{i_{n-k}}$  be the  $n-k$  radii that are lower or equal than the threshold  $u_{k,n} = R_{n-k,n}$ . Then  $A$  can be rewritten as

$$A = \frac{1}{n} \sum_{j=1}^{n-k} \mathbb{1}\{R_{i_j} \leq u_{k,n}\} \mathbb{1}\{\Theta_{\widehat{S}, i_j} \in \cdot\} = \left(1 - \frac{k}{n}\right) \frac{1}{n-k} \sum_{j=1}^{n-k} \mathbb{1}\{\Theta_{\widehat{S}, i_j} \in \cdot\}$$

with the latter sum being a sum of  $n-k$  iid terms conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$  as a consequence of the definition of  $\widehat{\Theta}_{i, u_{k,n}}^{(F)}$ ,  $i = 1, \dots, n$ . We will adapt the proof of Theorem 5.3 (ii) from Resnick (2006) (p. 138-139). Thus, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$

$$\frac{1}{n-k} \sum_{j=1}^{n-k} \mathbb{1}\{\Theta_{\widehat{S}, i_j} \in \cdot\} \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$  if and only if

$$\mathbb{E} \left( \frac{1}{n-k} \sum_{j=1}^{n-k} \mathbb{1}\{\Theta_{\widehat{S}, i_j} \in \cdot\} \middle| \mathbf{X}_1, \dots, \mathbf{X}_n \right) \xrightarrow{v} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$ . Now, we have

$$\mathbb{E} \left( \frac{1}{n-k} \sum_{j=1}^{n-k} \mathbb{1}\{\Theta_{\widehat{S}, i_j} \in \cdot\} \middle| \mathbf{X}_1, \dots, \mathbf{X}_n \right) = \mathbb{P}(\Theta_{\widehat{S}, 1} \in \cdot | \mathbf{X}_1, \dots, \mathbf{X}_n) = \widehat{S}(\cdot)$$

and

$$\widehat{S}(\cdot) \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$  (see Theorem 2). Hence, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$ ,  $A \xrightarrow{d} S(\cdot)$  in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

**Study of the term  $B$ :**

$$B = \frac{k}{n} \times \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{R_i > R_{n-k,n}, \Theta_i \in \cdot\} = \frac{k}{n} \widehat{S}(\cdot) \xrightarrow{\mathbb{P}} 0$$

since  $\widehat{S}(\cdot) \xrightarrow{d} S(\cdot)$  in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$  (see Theorem 2). We conclude that, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$

$$\widehat{S}_{u_{k,n}}^{(F)}(\cdot) \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

### 5.3 Proof of Proposition 2

As in the proof of Proposition 1, we have

$$\begin{aligned} \widehat{S}_{u_{k,n}}^{(F)}(\cdot) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\widehat{\Theta}_{i, u_{k,n}}^{(F)} \in \cdot\} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i \leq n/k, \Theta_{\widehat{S}, i} \in \cdot\} + \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i > n/k, \Theta_i \in \cdot\} \\ &=: C + D. \end{aligned}$$

We study each term separately.

**Study of the term C:**

Denote by  $\lfloor x \rfloor$  the integer part of  $x \geq 0$ . Using the same kind of arguments as in the proof of Theorem 2, we can show that

$$\frac{R_{n-\lfloor ck \rfloor, n}}{n/k} \xrightarrow{\mathbb{P}} 1$$

as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ . Moreover, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , we have

$$\frac{1}{n} \sum_{j=1}^{n-\lfloor ck \rfloor} \mathbb{1}\{\boldsymbol{\Theta}_{\widehat{S}, i_j} \in \cdot\} \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$ , as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$  (see the study of term A in the proof of Proposition 1). Then, as in the proof of Theorem 2, we prove by the continuous mapping theorem that, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$

$$C \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$ , as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

**Study of the term D:**

One can show that (see Resnick, 2006, Proposition 9.4, p. 312)

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\left\{\left(\frac{R_i}{n/k}, \boldsymbol{\Theta}_i\right) \in \cdot\right\} \xrightarrow{d} c\nu_1 \times S(\cdot) \quad (32)$$

in  $M_+(\mathbb{N}_+)$ , as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ . Then

$$D = \frac{k}{n} \times \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{R_i > n/k, \boldsymbol{\Theta}_i \in \cdot\} \xrightarrow{\mathbb{P}} 0.$$

Indeed, by (32), we have

$$\frac{1}{k} \sum_{i=1}^n \mathbb{1}\{R_i > n/k, \boldsymbol{\Theta}_i \in \cdot\} \xrightarrow{d} cS(\cdot)$$

in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

We conclude that, conditionally on  $\mathbf{X}_1, \dots, \mathbf{X}_n$

$$\widehat{S}_{u_{k,n}}^{(F)}(\cdot) \xrightarrow{d} S(\cdot)$$

in  $M_+(\mathbb{N}_+)$  as  $n \rightarrow \infty$ ,  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ .

## 6 Acknowledgements

Part of this work was supported by the FP7 ACQWA project, the GIS MOPERA project and by the ANR PEPER project. The authors are grateful to Johan Segers for providing the MELE R routine. The authors would like to thank an anonymous referee for valuable suggestions, which lead to an improved version of the article.

## References

BEIRLANT, J., DIERCKX G. & GUILLOU, A. (2005). Estimation of the extreme value index and regression on generalized quantile plots. *Bernoulli*, 11, 949-970.

- BEIRLANT, J., GOEGBEUR, Y., SEGERS, J. & TEUGELS, J.L. (2004). *Statistics of extremes: Theory and Applications*, John Wiley & Sons, Chichester.
- BEIRLANT, J., VYNCKIER, P. & TEUGELS, J.L. (1996). Excess functions and estimation of the extreme value index. *Bernoulli*, 2, 293-318.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed., New York, Wiley-Interscience.
- CEBRIÁN, A.C., DENUIT M. & LAMBERT, M. (2003). Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database. *Belgian Actuarial Bulletin*, 3:1, 33-41.
- CORCORAN, J.N. & SCHNEIDER, U. (2003). Shift and scale coupling methods for perfect simulation. *Probab. Engrg. Inform. Sci.*, 17, 277-303.
- DAVIS, R.A. & MIKOSCH, T. (1998). The sample autocorrelation of heavy-tailed processes with applications to arch. *Ann. Statist.*, 26, 2049-2080.
- DEKKERS, A.L.M., EINMAHL, J.H.J. & DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17, 1833-1855.
- EINMAHL, J.H.J., DE HAAN, L. & HUANG, X. (1993). Estimating a multidimensional extreme-value distribution. *J. Multivariate Anal.*, 47, 35-47.
- EINMAHL, J.H.J., DE HAAN, L. & PITERBARG, V.I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.*, 29, 1401-1423.
- EINMAHL, J.H.J., DE HAAN, L. & SINHA, A.K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.*, 70, 143-171.
- EINMAHL, J.H.J. & SEGERS, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution. *Ann. Statist.*, 37, 2953-2989.
- EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modelling extremal events for insurance and finance*, Springer-Verlag, Berlin.
- ENGLE, R.F. (1982). Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.
- FREES, E.W. & VALDEZ, E.A. (1998). Understanding using copulas. *N. Am. Actuar. J.*, 2, 1-25.
- GUMBEL, E.J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9, 171-173.
- DE HAAN, L. & FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*, Springer-Verlag, New York.
- DE HAAN, L. & RESNICK, S.I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40, 317-337.
- DE HAAN, L. & DE RONDE, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes*, 1, 7-46.
- HILL, B.M. (1975). A simple general approach about the tail of a distribution. *Ann. Statist.*, 3, 1163-1174.
- PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3, 119-131.

- RESNICK, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*, New York, Springer-Verlag.
- RESNICK, S.I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, New York, Springer-Verlag.
- TAWN, J.A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75, 397-415.
- WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the  $k$  largest observations. *J. Amer. Statist. Assoc.*, 73, 812-815.
- YOU, A., SCHNEIDER, U., GUILLOU, A., NAVEAU, P. (2010). Improving extreme quantile estimation via a folding procedure. *J. Statist. Plann. Inference*, 140, 1775-1787.
- Corresponding author:** Armelle Guillou, Université de Strasbourg et CNRS, IRMA UMR 7501, 7 Rue René Descartes, 67084 Strasbourg cedex, France, E-mail: armelle.guillou@math.unistra.fr.